

LA-4822

C. 1

The Notion of Complexity

For Reference

Not to be taken from this room



los alamos
scientific laboratory
of the University of California
LOS ALAMOS, NEW MEXICO 87544

This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Atomic Energy Commission, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe privately owned rights.

This report expresses the opinions of the author or authors and does not necessarily reflect the opinions or views of the Los Alamos Scientific Laboratory.

Printed in the United States of America. Available from
National Technical Information Service
U. S. Department of Commerce
5285 Port Royal Road
Springfield, Virginia 22151
Price: Printed Copy \$3.00; Microfiche \$0.95

LA-4822

UC-32

ISSUED: December 1971



Los Alamos
scientific laboratory
of the University of California
LOS ALAMOS, NEW MEXICO 87544

The Notion of Complexity

by

W. A. Beyer
M. L. Stein
S. M. Ulam



THE NOTION OF COMPLEXITY

by

W. A. Beyer, M. L. Stein, and S. M. Ulam

ABSTRACT

The notion of the arithmetic complexity $|n|$ of an integer n is defined in terms of the minimum number of additions, multiplications, and exponentiations required to combine 1's to form n . The value of $|n|$ is calculated for $n < 2^{10}$. n is called complicated if $|n| > |n_1|$ for every $n_1 < n$. Of the first 19 complicated numbers, 14 are prime. A conjecture about a relation between complexity and entropy is proposed. Some computations are presented to support this conjecture.

I. INTRODUCTION

In this report we discuss notions of complexity in some algebraic structures. These notions are also applicable to more general combinatorial situations that perhaps lack any algebraic pattern in the classical sense. We concentrate on a few special cases for which we studied and calculated a special notion of complexity. Essentially, we examined a special notion of complexity for ordinary integers with a little excursion on such a notion for integers modulo a prime.

The notion of complexity, in our view, is separate, though associated with the idea of the amount of information or entropy of a system. We mention briefly a possible axiomatic approach to defining a real number called complexity for elements of a set or of a class on which certain operations are performed. These could be binary operations; our set could be a set of integers, and the operations could be addition, multiplication, and exponentiation, for example. It is this case that was examined on a computing machine and to which most of this report is devoted.

Another case would be a class of subsets of a given set, with allowed operations being the Boolean operations of union and intersection or

union and complementation. One could add other operations, for example, the direct product of sets and also projection. This would correspond to allowing quantifiers in our theory. One can study a notion of complexity for vectors in a countable space or even in the continuum. An important study would be that of a relative complexity; that is to say, complexity of elements or "expressions" when the complexity of certain symbols is normalized to 1. In what has been sometimes called "speculation" on constants in physical theories, for example, the whole art seems to depend on the success of attempts to define some known important numbers, e.g., the dimensionless ratios

$$M_{\text{proton}}/M_{\text{electron}} = 1836.11\dots$$

and

$$e^2/hc = 137.1\dots$$

by use of only a few artificially introduced constants which should be as "simple" as possible. (cf. the attempts by Eddington¹ and some very recent ones by Good² and Wyler.³)

Considered "genetically," a mathematical theory resembles a tree in that one obtains from a given number of symbols corresponding to "variables"

and from a number of allowed operations, expressions that elongate by branching. The simplifications and abbreviations may then reduce the length of the expressions.

One could try to define complexity in a mathematical structure by postulating certain of its properties, somewhat like postulating properties of a measure.

Let the structure, S , consist of elements x , y , ... It may be finite or infinite. We have in the set S a number of, say, binary operations R_1 , R_2 , ... R_n . We want to assign a number $c(x) \geq 0$ to each element x of S and to each R_i ($i = 1 \dots n$) so that the following properties should hold.

- a. If $z = R_i(x, y)$, then $c(z) = c(R_i(x, y)) \leq c(x) + c(y) + c(R_i)$ $i = 1 \dots n$.
- b. For each element z , if $z = R_j(x, y)$, we should have for one case at least, $c(z) = c(x) + c(y) + c(R_j)$.
- c. $z(x_0) = z(x_1) = \dots = z(x_n)$ for some pre-assigned elements $x_0 \dots x_n$.

Needless to say, one can define analogous desiderata for the case in which the operations are more general than binary ones.

Obviously, in the case to which our exercise is devoted, these postulates are satisfied. Moreover, they define the complexity uniquely if, as must be the case in general, the complexity was normalized for some elements. (In our case, we assume the complexity of the integer 0 to be equal to 1. We hope to study this notion more thoroughly for the more general case and also to perform experiments to determine complexity functions for the case in which S is a class of sets.) Ultimately, one would wish to discuss the complexity of genetic codes and biological organisms quantitatively.

("Integer" always means a positive integer.)

II. ARITHMETIC COMPLEXITY OF INTEGERS

The arithmetic complexity $|n|$ of an integer n is defined as the fewest number of operators: $+$, \times , $\times\times$ (addition, multiplication, and exponentiation) which combine 1's to form n . Thus, $|1| = 0$; $|2| = 1$ since $2 = 1 + 1$; and $|5| = 4$ since $5 = (1 + 1)\times\times(1 + 1) + 1$ and not fewer than four operators with 1's will form five. Obviously, for a and b integers, $|a + b|$, $|ab|$, and $|a^b|$ are each not more than

$|a| + |b| + 1$. For an infinity of integers n , the relation $|n + 1| = |n| + 1$ holds.

For the purpose of calculating the complexity of some integers, all correct formulas (up to some number of operators) involving $+$, \times , $\times\times$, and the number 1 were enumerated using parenthesis-free notation on a computer. It required one hour of computer time to enumerate the integers with complexity ≤ 6 . Ralph Cooper made the following observation. Each correct formula involving n (> 0) operators is the composition of two formulas, one formula with n_1 operators and one formula with n_2 operators such that $n = n_1 + n_2 + 1$. One generates the integers of complexity n by first generating tables of integers of complexity $< n$. One partitions $n - 1$ into $n_1 + n_2$ in all ways and combines the integers of complexity n_1 with the integers of complexity n_2 to produce integers of complexity not larger than n . This method is considerably more efficient than the previous method. Table I lists the complexity of all integers $< 2^{10}$.

From the above construction, one sees that an upper bound $\ell^1(k)$ to $\ell(k)$, the number of integers of complexity k , is given by the solution of

$$\ell^1(k + 1) = \sum_{j=0}^k \ell^1(j) \ell^1(k - j),$$

with $\ell^1(0) = 1$. The solution to this equation is given by

$$\ell^1(k) = \frac{1}{k+1} \binom{2k}{k} 2^{-k},$$

which implies that

$$\ell(k) \leq \frac{2^k}{k\sqrt{\pi k}} + o\left(2^k k^{-5/2}\right).$$

Two additional forms of complexity have been considered and calculated.

- a. Complement complexity. To make complexity symmetric in 0's and 1's, we introduce a slightly different complexity, the complement complexity $\bar{K}(y|n)$. Define the complement operation C by $C(x|n) = 2^n - 1 - x$. $\bar{K}(y|n)$ is defined as the fewest operations of addition, multiplication, exponentiation, and complementation that combine 1's to form y . In the count of operations, the

TABLE I. COMPLEXITY OF INTEGERS $< 2^{10}$.

Complexity	Integer
0	1
1	2
2	3
3	4
4	5 6 8 9
5	7 10 16 27
6	11 12 17 18 25 28 32 36 64 81 256 512
7	13 14 15 19 20 24 26 29 33 37 49 54 65 82 100 125 128 216 243 257 513 729 1024
8	21 22 30 34 38 48 50 55 56 66 72 83 101 121 126 129 144 162 217 244 256 289 324 343 514 625 730 784 1000
9	23 31 35 39 40 45 51 52 57 58 67 73 74 75 84 96 98 102 108 122 127 130 145 163 164 169 192 196 200 218 225 245 250 259 290 325 344 361 400 432 486 515 576 626 676 731 768 785 844 1001
10	41 42 44 46 53 59 60 63 68 76 78 80 85 87 90 97 99 103 109 110 111 112 123 131 132 135 146 147 165 166 170 193 195 197 201 202 219 226 242 246 251 252 260 286 291 300 326 345 362 375 384 401 433 434 441 484 487 488 516 577 578 627 646 677 686 732 769 771 784 842 900 1002
11	43 47 61 62 69 70 77 79 86 88 89 91 104 113 114 116 124 133 134 136 140 146 150 153 160 167 168 171 180 189 194 198 203 204 220 224 227 247 249 253 254 261 262 264 265 270 292 301 303 320 327 328 336 346 363 376 378 385 387 392 402 405 435 436 442 450 485 489 490 500 517 518 520 521 529 579 580 626 649 650 651 678 687 688 722 733 770 772 774 787 800 843 864 867 901 961 972 1003
12	71 92 93 95 105 106 115 117 118 119 120 137 141 149 151 152 154 156 161 172 174 175 176 181 185 190 199 205 206 208 221 222 226 232 234 248 255 263 266 271 272 280 263 293 294 296 297 302 304 306 321 329 330 332 333 339 340 347 360 364 366 377 379 381 386 388 390 393 394 403 404 406 410 437 438 443 446 451 452 459 491 492 501 502 504 507 519 522 528 530 539 567 581 582 585 588 600 629 640 652 654 656 675 679 689 690 723 724 734 735 737 738 750 756 773 775 777 788 801 802 810 844 865 866 868 870 882 902 962 968 973 974 975 976 1004
13	94 107 138 142 155 157 158 159 173 177 178 182 186 187 191 207 209 223 229 231 233 235 240 247 268 273 274 275 281 284 295 298 305 307 308 309 322 331 334 336 337 341 342 346 349 351 352 365 367 369 370 380 382 389 391 395 396 407 408 411 415 416 425 439 440 444 449 453 454 455 460 464 476 493 494 495 498 503 505 506 504 510 523 524 531 537 540 544 548 568 574 583 584 586 589 591 592 593 594 601 602 603 605 606 612 630 631 633 634 641 645 653 655 657 664 680 691 692 700 702 704 720 725 726 736 739 745 747 751 752 753 757 776 778 780 783 789 790 792 793 803 804 806 811 820 845 869 871 872 873 875 883 884 891 896 903 909 963 969 970 977 978 960 999 1005 1006 1008 1009
14	139 143 179 183 184 188 210 212 230 236 237 238 241 269 276 279 282 285 286 299 310 312 315 316 319 323 335 350 353 356 359 368 371 372 383 397 398 399 409 412 417 420 424 445 456 461 462 465 468 472 475 477 480 496 499 509 511 525 526 527 532 536 538 541 542 545 549 550 560 561 566 569 575 587 590 595 604 607 608 609 610 613 632 635 637 642 646 658 660 665 666 672 681 682 684 685 693 694 701 703 705 707 715 721 727 728 740 741 746 748 754 755 758 759 761 762 765 779 781 791 794 795 805 806 809 812 815 816 821 825 830 832 833 846 847 849 850 874 876 880 885 886 892 897 904 910 918 924 925 928 936 960 964 971 979 981 982 984 985 1007 1010 1014 1016
15	211 213 214 239 277 278 287 311 313 314 317 318 354 355 357 373 374 413 414 418 421 423 424 427 429 446 447 457 459 463 466 469 470 473 478 481 483 497 533 534 543 546 541 555 562 570 596 597 599 611 614 615 616 618 621 624 636 638 643 644 647 659 661 662 663 667 668 670 673 674 683 695 696 698 706 708 714 716 742 743 744 749 760 763 764 766 782 796 798 807 813 814 817 822 824 826 829 831 834 836 837 840 848 851 854 855 857 858 877 878 879 881 887 888 889 893 898 905 906 908 911 912 913 919 920 926 927 929 931 935 937 945 950 952 957 965 968 986 987 988 990 996 1011 1012 1015 1017 1018 1020
16	215 358 419 422 428 430 467 471 474 479 482 535 547 552 556 557 558 559 563 564 565 571 572 573 594 617 619 620 622 639 669 671 697 699 709 711 712 713 717 718 767 797 799 818 819 823 827 828 835 838 852 853 856 859 861 890 894 899 907 914 915 916 917 921 922 930 932 938 944 946 951 953 954 958 966 967 989 991 992 993 997 998 1013 1019 1021 1022 1023
17	431 553 554 623 710 719 839 860 862 895 923 933 939 940 941 942 947 948 949 955 956 959 994 995

first three are given the value 1 and the last is given the value zero. Thus $\bar{K}(y|n) = \bar{K}(2^n - 1 - y|n)$. Table II gives the values of $\bar{K}(y|n)$ for $y < 2^{10}$ and $n = 10$.

- b. Modulo a prime p complexity. In addition to the operations of +, x, and xx, the operation of mod_p is allowed and is defined by $\text{mod}_p(x) = x - p[x/p]$ where p is a fixed prime and $[]$ denotes the greatest integer. Table III gives the modulo prime $p = 137$ complexity for integers < 137 . Table IV gives the modulo prime $p = 1009$ complexity for integers < 1009 .

III. COMPLICATED NUMBERS

One defines n to be a complicated number if $|n| > |n_1|$ for every $n_1 < n$. The complicated numbers $< 2^{10}$ are 1, 2, 3, 4, 5, 7, 11, 13, 21, 23, 41, 43, 71, 94, 139, 211, 215, 431, and 863. (Those underlined are also prime.) Obviously, there are an infinity of complicated numbers. We propose the following conjectures.

- There exists K such that all complicated numbers $K_1 > K$ are prime.
- Every sufficiently large integer n is the sum of $k < \log n$ complicated integers.
- There exists c such that every sufficiently large n satisfies $|n| < c + \sqrt{\log n}$.

IV. COMPLEXITY AND ENTROPY

Kolmogorov^{4,5} has introduced the notion of complexity of a finite string over a given alphabet. For simplicity, suppose the alphabet to be $\{0,1\}$. Let A be an algorithm that transforms finite binary sequences into binary sequences. By an algorithm is meant any of the various equivalent concepts used in logic. For a binary string x, one defines the complexity by

$$K_A(x) = \begin{cases} \min \ell(p) \\ A(p)=x \\ \infty \\ \text{if no } p \text{ exists such that } A(p) = x, \end{cases}$$

where $\ell(p)$ denotes the length of the binary string p. Analogously, one defines conditional complexity.

Let $A(p,x)$ be an algorithm defined from pairs of binary strings to binary strings. Put

$$K_A(y|x) = \begin{cases} \min \ell(p) \\ A(p,x)=y \\ \infty \\ \text{if no } p \text{ exists such that } A(p,x) = y. \end{cases}$$

$K_A(y|x)$ is called the conditional complexity of y with respect to x. Kolmogorov regards complexity as analogous to entropy. We make the following conjecture.

Conjecture. Let a discrete binary information source S in the sense of Shannon⁶ be given with entropy $H = -p \log p - (1-p) \log (1-p)$ where probability (0) = p and probability (1) = 1-p; $0 < p < 1$. Let $\{x_1, x_2, \dots, x_{2^n}\}$ be the set of all binary strings of length n arranged in order of decreasing probability. Let $k(n)$ be the least integer so that $\sum_{i=1}^{k(n)} \text{prob}(x_i) > r$ where $1/2 < r < 1$. Then asymptotically for large n,

$$H \approx \frac{1}{k(n)} \sum_{i=1}^{k(n)} K_A(x_i|n). \quad (1)$$

(In Eq. (1), K_A should be normalized so that when $p = 1/2$,

$$\frac{1}{k(n)} \sum_{i=1}^{k(n)} K_A(x_i|n) = 1.)$$

In other words, the most likely sequences from A have complexity approximately equal to the entropy of S.

In order to test the conjecture expressed in Eq. (1), we replaced $K_A(x_i|n)$ by $\lambda \bar{K}(y|n)$, where λ is selected so that when $p = 1/2$,

$$\frac{1}{k(n)} \sum_{i=1}^{k(n)} \lambda \bar{K}(x_i|n) = 1.$$

Graphs of $H_1 = -p \log p - (1-p) \log (1-p)$ and

$$H_2 = \frac{1}{k(n)} \sum_{i=1}^n \lambda \bar{K}(x_i|n)$$

when $n = 10$ and $r = .75$ are shown in Fig. 1

TABLE III. MODULO PRIME $p = 137$ COMPLEXITY OF INTEGERS < 137 .

Complexity	Integer																			
0	1																			
1	2																			
2	3																			
3	4																			
4	5	6	8	9																
5	7	10	14	27																
6	11	12	17	16	25	24	32	36	64	81	101	119								
7	13	14	15	19	20	24	26	29	33	37	44	49	50	54	61	65	79	82	92	100
8	21	22	30	34	38	41	45	48	51	55	56	60	62	63	66	68	69	72	77	80
9	23	31	35	39	40	42	46	47	52	53	57	58	59	67	70	73	74	75	76	78
10	0	43	71	85	86	90	95	97	105	114	116	135								
11	91																			

V. COMPLEXITY OF N-TUPLES OF INTEGERS

Matijasevič⁷ has proved the following theorem. There exists a fifth-degree polynomial $Q(y_1, \dots, y_k; z)$ with integer coefficients such that any enumerable set m of natural numbers (for example, the set of prime numbers) coincides with the set of natural values of the polynomial $Q(y_1, \dots, y_k; a_m)$ where a_m is a certain number effectively constructed for the set m . From the result, it follows that if one could discuss complexity of n -tuples of integers, then one could discuss the complexity of enumerable sets of natural numbers by equating such complexity to the complexity of the associated polynomial Q .

REFERENCES

1. A. S. Eddington, Fundamental Theory (Cambridge University Press, 1946).
2. I. J. Good, "The Proton and Neutron Masses and a Conjecture for the Gravitational Constant," Phys. Let. **33A**, 383-384 (1970).
3. A. Wyler, "Les Groupes des Potentiels de Coulomb et de Yukawa," Compt. Rend. Acad. Sci. Paris, **271**, 186-188 (1971).
4. A. Kolmogorov, "Three Approaches for Defining the Concept of Information Quantity," Problems of Information Transmission **1**, 3-11 (1965).

5. A. Kolmogorov, "Logical Basis for Information Theory and Probability Theory," IEEE Trans. Information Theory **IT-14**, 662-664 (1968).
6. C. E. Shannon and Warren Weaver, The Mathematical Theory of Communication (The University of Illinois Press, Urbana, 1949).
7. Ju V. Matijasevič, "Enumerable Sets are Diophantine," Soviet Math. Dokl. **11**, 354-358 (1970).

ADDITIONAL REFERENCES TO COMPLEXITY NOT USED IN TEXT

1. E. L. Lawler, "The Complexity of Combinatorial Computations: A Survey," Proceedings of 1971 Polytechnic Institute of Brooklyn Symposium on Computers and Automata.
2. D. W. Loveland, "On Minimal Program Complexity Measures," ACM Symp. Theory of Computing, Marina del Rey, California, May 5-7, 1969.
3. P. Young, "A Note on Dense and Nondense Families of Complexity Classes," Math. Systems Theory **5**, 66-70 (1971).

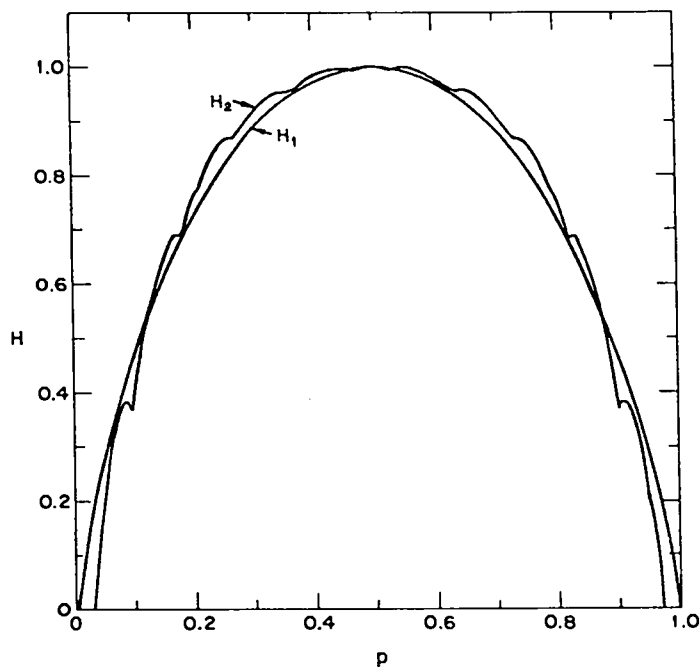


Fig. 1. Comparison of entropy $H_1 = - \sum p_i \log p_i$ and complement complexity H_2 as defined and discussed in text.